Similarity Effects on Recognition: Testing the Diagnostic Feature-Detection Hypothesis

Ellie Newton

Word Count: 9986

A report submitted as partial requirement for the degree of Bachelor of Psychological Science with Honours in Psychology at the University of Tasmania, 2019.

Statement of Sources

I declare that this report is my own original work that contributions of others have been duly acknowledged: Ellie Newton, 17/9/2019.

Acknowledgments

First and foremost, I would like to thank my supervisor Dr Jim Sauer for his guidance throughout the year. I am especially grateful for his continuous support, expertise and excitement for science during every step in the process of completing my thesis.

I would also like to thank Matthew Gretton for designing the software for my study and for taking the time to explain the complex aspects of this to me. I would like to extend my thanks to Dr Matt Palmer for sharing his expertise in this area throughout the year. Finally, I would like to thank everyone in the Cognition Lab, for their constant support, advice and chocolate.

Table of Contents

Abstract 1

Introduction 2

Recognition Decision-Making 4

Lineup Procedures and Underlying Mechanisms 5

Measures of Performance: Diagnosticity vs. ROC Analysis 6

Diagnostic Feature-Detection Theory 9

Similarity Effects on Facial Recognition 14

The Current Study 18

Hypotheses 20

Method 21

Design 21

Participants 22

Materials 22

Stimuli 22

Procedure 24

Results 26

Statistical Strategy 26

Testing the DFD 28

Exploratory Analysis 30

Discussion 30

Tests of the DFD and Relative Familiarity Mechanism 31

Implications for Lineup Construction 34

Limitations and Future Directions 34

Conclusions 37

References 39

Appendix A 45

Appendix B 50

Appendix C 52

List of Tables and Figures

Figure 1.....8

Figure 2.....21

Figure 3.....23

Figure 4.....23

Table 1.....29

Figure 5.....29

Table 2.....45

Table 3.....46

Table 4.....48

Table 5.....49

Abstract

Diagnostic feature-detection (DFD) hypothesis proposes that the simultaneous presentation of faces (e.g., in police lineups) enhances identification accuracy by allowing shared features to be discounted and diagnostic features to be detected. However, no direct tests of this theory have been conducted in face recognition tasks. We directly tested the DFD hypothesis. 60 participants (42 women; aged 18 to 45 years) completed a facial recognition task. Memory for faces was tested in three task conditions: a yes/no (Y/N) task, in which one face was presented at a time; a two alternative forced choice (2AFC) task, in which participants had to choose one of two faces, presented simultaneously; a two alternative open choice (2AOC) task, in which two faces were presented simultaneously, with a 'not present' option. Difficulty was manipulated through two levels of target-lure similarity (high, low). Discrimination (*d'*) was calculated and compared between the high and low similarity trials within each condition. Contrary to the predictions of DFD, we found significant differences in discrimination between high and low similarity trials in the 2AFC and 2AOC conditions. Unexpectedly, the manipulation of difficulty did not affect performance in the Y/N condition. We found no support for the DFD hypothesis.

After witnessing a crime, individuals are often faced with an identification task (Sauer, Palmer & Brewer, 2019). This often takes the form of a lineup, which places one suspect (innocent or guilty) amongst several known-to-be-innocent fillers. The witness then indicates who, if anyone, they recognise as the culprit. Their response will have ramifications for the investigation. A rejection of the lineup may deter investigations away from the suspect. Identification of the suspect encourages further investigation and increases the likelihood of prosecution. An eyewitness identification is an especially convincing piece of evidence for jurors. An identification appears clear-cut amongst the other, often ambiguous evidence decision-makers in the criminal justice system must deal with and, for this reason, identifications are often given considerable weight. Often, at least early in an investigation, an identification is the primary (or only) piece of evidence investigators have to work with (Cutler & Penrod, 1995). Further, sometimes identifications are the chief (or even only) piece of evidence supporting a prosecution. Eyewitness identifications are common and compelling, and thus important to understand (Brewer, Caon, Todd & Weber, 2006; Steblay, Dysart, Fulero & Lindsay, 2001; Steblay, Dysart, & Wells, 2011).

Given their reliance on human cognition, primarily memory, identifications are unsurprisingly error-prone. Evidence of eyewitness error comes from both controlled lab experiments and real-world cases. According to the Innocence Project (2019), 70% of wrongful convictions exonerated with DNA evidence were attributable, at least partly, to false identifications. Awareness of the contribution of identification error to wrongful convictions has prompted researchers to investigate the conditions under which identification errors are more likely, and procedural innovations that can reduce risk of error. Two key procedural issues identified in the

identification literature are lineup composition and presentation method, which essentially ask: who should we be putting in the lineup to make it fair, (how similar should fillers be to the suspect, and on what criteria should similarity be judged) and how do we display that lineup to maximize performance (Fitzgerald, Price, Oriet & Charman, 2013).

Lineup presentation is a contentious issue. Should eyewitnesses view all lineup members at once—in what is called a simultaneous lineup—or should they view each lineup member in isolation—in a sequential lineup. The literature has suggested that sequential (cf. simultaneous) presentation improves performance by reducing false identifications of innocent suspects (Steblay et al., 2001; Steblay et al., 2011). However, some researchers have recently found evidence of improved performance for simultaneous (cf. sequential) presentation (Mickes, Flowe & Wixted, 2012). In particular, simultaneous presentation appears especially useful when the innocent suspect is highly plausible, or when innocent fillers are highly similar to each other and the suspect. The diagnostic feature-detection hypothesis has been put forward as the explanation for this, which argues that simultaneous presentation allows the eyewitness to detect and discount shared features, and better detect features that are diagnostic of guilt (Wixted & Mickes, 2014). Although researchers have presented data consistent with this mechanism, it has not been directly tested.

In the current research, we directly tested whether simultaneous presentation and the diagnostic feature-detection mechanisms attenuated the deleterious effects of increased suspect-filler similarity on identification performance. Suspect-filler similarity was manipulated to create two levels of difficulty. In general recognition tasks, highly plausible (i.e., similar) lures/fillers increase task difficultly (i.e., it is

harder to choose correctly). In lineups, the task is made more difficult when an innocent suspect is a highly plausible choice, or when the fillers are highly similar to the culprit. However, diagnostic feature-detection theory holds that simultaneous presentation increases the witness's ability to distinguish between a face they have seen before and one they have not, attenuating these similarity effects (Wixted & Mickes, 2014).

Recognition Decision-Making

In various decision-making tasks (e.g., perceptual discrimination and recognition memory), researchers often adopt a signal detection theory (SDT) framework when investigating effects on performance (Macmillan & Creelman, 2004; Wixted & Mickes, 2014). SDT teases apart the two key mechanisms of decision performance: discriminability and response bias (Green & Swets, 1966). Discriminability refers to the ability to discriminate between a previously seen target and an unseen lure, or in an identification, between a previously seen (i.e., guilty) and unseen (innocent) lineup member. Response bias refers to an overall tendency/willingness to classify a test item as previously seen. Response bias indicates where an individual's decision criterion is placed along a continuum of evidence, and can vary from extremely lenient to extremely conservative. A positive recognition decision (or identification) is made if the strength of the individual's memory signal exceeds this criterion (Wixted & Mickes, 2014). In the identification context, a memory signal refers to the sense of "familiarity" an individual experiences when comparing a lineup member (i.e., the combination of that lineup member's facial features) to their memory of the culprit (Colloff, Wade & Strange, 2016). The eyewitness's decision will therefore reflect both memorial factors and external factors (e.g., expectations and administrator instructions) operating on their

response criterion. However, an identification will ideally index a comparison of a lineup member with the witness's memory of the culprit. Thus, the memory signal (rather than response bias) would ideally drive of the decision.

Lineup Procedures and Underlying Mechanisms

Researchers have generally focused on identifying the presentation method that maximises overall identification accuracy. There are three common identification test procedures. Show ups presents a single photo to the eyewitness as an old or new recognition task (Lindsay & Wells, 1985). However, with only one face presented, this procedure is highly suggestive of suspect guilt and, therefore, is generally frowned upon (Wixted & Mickes, 2014). The two procedures that have received the most attention in the literature are both lineup formats, which involve placing a suspect amongst innocent fillers (Colloff & Wixted, 2019; Wixted, Vul, Mickes & Wilson, 2018; Sauer et al., 2019).

In a simultaneous lineup, all lineup members are presented together. Simultaneous lineups promote a relative decision strategy, whereby lineup members are compared to each other for *relative* similarity to the eyewitness' memory of the culprit (Steblay, Dysart, & Wells, 2011). Relative familiarity can work well when the culprit is present to be recognised, but leads to problems if the culprit is absent. An identification decision should indicate that the identified lineup member provides a strong match to the witness's memory, not simply that that the identified lineup was the best available option. Thus, ideally eyewitnesses will compare each lineup member directly to their memory of the culprit, thereby facilitating an *absolute judgement* – a decision based on the absolute degree to which a lineup member matches the witness's memory of the culprit. The absolute and relative judgement distinction has been the dominant theoretical framework for understanding how

presentation methods affects eyewitness recognition decisions. The sequential lineup was developed to reduce the likelihood of misidentifications, by discouraging use of relative judgements and encouraging the witness to make an absolute judgement (Lindsay & Wells, 1985; Steblay et al., 2011). Sequential lineups present lineup members one at time, and each face requires a 'yes' or 'no' response before the next lineup member can be viewed. Once an identification ('yes' response) has been made, the procedure is over (Brewer & Wells, 2006; Lindsay & Wells, 1985).

Meta-analyses suggest that sequential lineups produce similar correct identification rates as simultaneous lineups while reducing false identifications (Steblay et al., 2001). For this reason, sequential lineups have been strongly recommended by prominent eyewitness researchers (e.g., Wells, 2014). However, Wixted and Mickes (2014) argued that, in SDT terms, these differences reflect variations in response bias. That is, sequential presentation does not make witnesses *better*, it just makes them more conservative. Wixted and Mickes (2014) argue that a more conservative criterion placement explains the lower rates of choosing, and thus the lower rates of false identifications observed in sequential lineups. Further, Palmer & Brewer's (2012) meta-analysis found that sequential presentation did not affect discriminability, but produced a conservative shift in participants' decision criteria. Wixted & Mickes (2014) maintain that the superior lineup procedure should improve performance through enhanced discriminability, not response bias.

Measures of Performance: Diagnosticity vs. ROC Analysis

Previously, researchers evaluated lineup procedures using diagnosticity ratios. In eyewitness research, the proportion of correct identifications of the culprit in target present trials is referred to as the hit rate (HR). The proportion of false identifications of the innocent suspect from target absent lineups is referred to as the

false alarm rate (FAR; Mickes et al., 2012). Diagnosticity ratios are calculated by dividing the HR by the FAR and tell us, given the suspect was identified, how likely is it that the suspect is guilty. Comparing diagnosticity ratios tells us which procedure produces identifications with the greatest diagnostic value. Steblay et al.'s (2011) meta-analysis found that sequential presentation produced higher diagnosticity ratios than simultaneous presentation. Although the simultaneous lineup had the higher HR, it also had the higher FAR (Mickes et al., 2012). Given that western criminal justice systems prioritise protecting the innocent, it is understandable that researchers would favour procedures that reduce false identifications, despite the cost of missing some guilty suspects. However, Wixted and Mickes (2014) criticised diagnosticity ratios, arguing that they measure response bias rather than discriminability. A higher diagnosticity ratio is associated with lower overall choosing rates, indicating a more conservative criterion (Mickes et al., 2012). Although reducing choosing rates reduces misidentification rates, it also reduces correct identification rates. Therefore, according to these researchers, diagnosticity ratios are confounded by choosing rates and unsuitable for investigating the superiority of lineup procedures.

Receiver Operating Characteristic (ROC) analysis is grounded in signal detection theory, and frequently used in the medical literature (Mickes et al., 2012). ROC analysis is based on multiple HR-FAR pairs, called operating points, which are plotted to form a ROC curve (see *Figure 1* below for an example). For a single diagnostic test, multiple pairs are obtained by varying the cut-off of a range of scores. In eyewitness literature, cut-offs are often based on confidence ratings attained from participants following identification decisions. These can vary from liberal to conservative cut-offs (e.g. decisions associated with >30% confidence

compared to >90%). The diagnostic value of a test (or lineup) is indexed by the area between the curve and the diagonal line (AUC). The more diagnostic a test, the more the ROC will bow toward to the top left corner, and the greater the AUC will be. The diagonal line on a ROC analysis demonstrates performance that provides no diagnostic information (i.e., HR = FAR). Mickes et al. (2012) described why ROC analysis is the preferable measure of diagnosticity accuracy compared to diagnosticity ratios. According to Mickes et al., diagnosticity ratios are based on a single HR-FAR pair and a single pair is incapable of characterising the performance of a diagnostic test, as that single pair could have a number of different ROC curves drawn through it. For this reason, comparing the performance of two lineup procedures with their diagnosticity ratios (i.e., their single HR-FAR pair) is insufficient (Mickes et al., 2012).

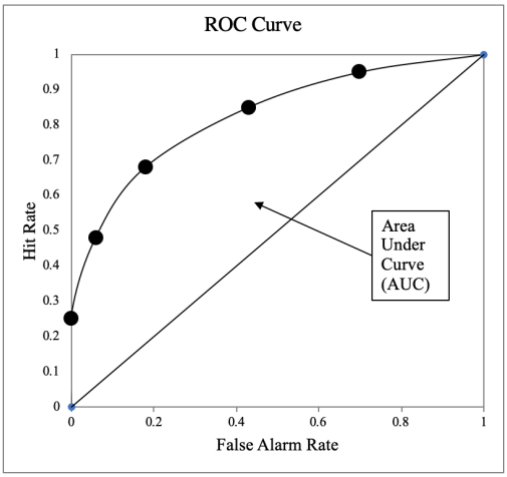

Figure 1. Hypothetical ROC curve plotted with multiple HR-FAR pairs.

Mickes et al. (2012) compared the ROC curves produced by witnesses making identifications from sequential and simultaneous procedures. Their analysis showed no evidence for a sequential superiority effect. In fact, they found a simultaneous superiority effect. Similar findings have been reported by Gronlund, Carlson, Dailey, and Goodsell (2009), Gronlund et al. (2012), and Wixted, Mickes,

Dunn, Clark and Wells (2016). For this reason, Wixted and Mickes (2014) called for greater research focus on discriminability and proposed a signal-detection-based model of eyewitness identification, measuring discriminability with ROC analysis. Mickes et al. (2012) confirmed that sequential lineups have a higher diagnosticity ratio, which infers that an identification from a sequential lineup has a higher probative value (i.e., more likely to be accurate). However, the sequential lineup had a 32% lower HR compared to the simultaneous, and simultaneous presentation produced a higher level of discriminability (Mickes et al., 2012; Wixted & Mickes, 2014). The diagnostic feature-detection hypothesis was presented as an explanation for these findings.

Diagnostic Feature-Detection Theory

Diagnostic feature-detection (DFD) theory proposes that performance in discrimination tasks is improved when the stimuli (e.g., lineup members) are presented simultaneously. The theory was developed with basic discrimination tasks (e.g., where participants determine which of two lines is longer), and accounts for the intuitive finding that such tasks are easier if the lines are presented together (because their length can be directly compared) than when lines are presented one after the other. More recently, DFD theory has been extended to account for improved discrimination of similar faces when presented simultaneously (Mundy, Honey, & Dwyer, 2007). Extending this to lineups, presenting a lineup simultaneously may allow for features shared by individuals to be recognised as non-diagnostic of guilt, and features that stand out to be recognised as more diagnostic for discriminating between innocent and guilty lineup members (Wixted & Mickes, 2014). For example, imagine an innocent suspect who shares some features (e.g., hair colour and style, skin colour, eye colour, etc.) with the culprit. If this suspect is presented to

the witness in isolation (i.e., in a showup), these shared features might provide a strong enough sense of familiarity for the witness to "recognise" and identify the innocent suspect. Whereas if the innocent suspect is presented in a simultaneous lineup with other lineup members who, importantly, share these features, the witness can discount theseshared features as no longer diagnostic, and search for non-shared diagnostic features to support their identification. According to DFD, simultaneous presentation enhances discriminability by helping witnesses to detect diagnostic features. The predictions of the DFD hypothesis are clear in some instances, but less clear in others. One clear prediction from DFD concerns the process of an identification from a simultaneous target present lineup (described above). But DFD-based hypotheses for target absent lineups are ambiguous: it is unclear whether DFD predicts that witnesses should be better able to correctly reject a simultaneous (cf. sequential) lineup. The DFD hypothesis has been proposed to explain demonstrations of improved discriminability in simultaneous lineups, but few direct tests of this theory *as an account of the mechanism* underlying performance differences between sequential and simultaneous lineups have been carried out.

For example, Colloff et al. (2016) investigated whether unfair lineups impair witnesses' ability to discriminate between innocent and guilty culprits. In an unfair lineup, the suspect stands out due to distinctive features (e.g., by possessing a distinctive tattoo or a black eye). In cases like these, lineups can be made fair by ensuring the eyewitness is unable to rely solely on this distinctive feature to make the identification. Building on a technique proposed by Zarkadi, Wade, and Stewart (2009), fair lineups were created with three methods: (1) replicating the feature; (2) pixelating the feature; (3) blacking out the feature, on all lineup members. These methods were compared to an unfair lineup, in which nothing was done to cover the

distinctive feature. Allowing the distinctive features to stand out on the suspect in a six-person lineup, increased participants' willingness to choose the suspect, but reduced their ability to discriminate between innocent and guilty suspects. Specifically, in the unfair condition, participants correctly identified the guilty suspect from target present lineups 57% of the time, but in target absent lineups, misidentified the innocent suspect 36% of the time. In comparison, the innocent suspect was misidentified less than 10% of the time in the three fair lineups. The fair lineups were associated with increased filler identifications from both target present and absent lineups, resulting in less correct identifications. The authors concluded that consistent with DFD predictions, allowing a distinctive/diagnostic feature to stand out will increase witnesses' willingness to choose the suspect. While the fair lineups protected innocent suspects from being falsely identified, they also led to great filler identifications and reduced rates of correct identifications of the guilty suspect. The authors have focused on the protective element of the fair lineups, while ignoring the reduced discriminability – which is not what DFD would have predicted due to the use of simultaneous presentation. Some of Colloff's et al. (2016) findings are potentially consistent with DFD predictions, but they did not directly test DFD as an explanation for differences in discrimination observed between sequential and simultaneous (they included only simultaneous lineups).

Further, Smith, Wells, Smalarz & Lampinen (2018) disputed Colloff's et al. (2016) conclusion, arguing that the high-similarity lineups did not improve performance in the study. Smith et al. (2018) relabelled the fair conditions as high similarity and the unfair condition as low similarity. Smith et al. presented differential filler siphoning as the more appropriate explanation for Colloff's et al. findings. This explanation suggests that when fillers are highly similar, witnesses are

less likely to identify the suspect, and false positives are siphoned away from the innocent suspect onto fillers. It is differential in that filler siphoning is more likely to occur if the suspect is innocent, rather than guilty. Smith et al. (2018) also argued that Colloff et al.'s participants spread their identifications across fillers (cf. correctly rejecting the lineup), and therefore concluded that increased similarity did not improve discriminability. With this point, Smith et al. (2018) have identified a key flaw in DFD theory. This theory does not make clear predictions concerning how a witness correctly rejects a target absent lineup. Although DFD describes how a witness identifies the culprit when he/she is there to be recognised it does not discuss the process of recognising that *none* of the lineup members are the culprit and deciding to reject the lineup. This forms a key part of our rationale for including target absent trials in our direct test of DFD theory as an account of performance differences observed in face recognition tasks using sequential vs. simultaneous arrays.

Colloff, Wade, Strange and Wixted (2018) agreed with Smith et al. (2018) that their results were consistent with differential filler siphoning. Specifically, that filler siphoning occurs more frequently in fair lineups (cf. unfair lineups).However, at the same time, they argued that filler siphoning is simply analogous to response bias, whereby responding becomes more conservative as the lineup becomes increasingly fair. Colloff et al. maintained that DFD was the more appropriate explanation for their previous findings because, unlike filler siphoning, DFD specifically predicts increased discrimination as the lineup becomes increasingly fair. Colloff et al. (2018) conducted another study employing a show up procedure to eliminate the possibility of a filler siphoning explanation. However, this also eliminated DFD as a possible explanation. Colloff et al. claimed that, according to

DFD, by using a fair showup (i.e., suspect does not have a distinctive feature), the witness is unable to rely on the diagnostic feature for identification and this will increase discrimination. However, DFD theory is based on the idea that simultaneous presentation is what allows eyewitnesses to discount shared features from *two or more* faces (or other stimuli) and this is what enhances discrimination (Wixted & Mickes, 2014). Colloff et al.'s application of DFD here demonstrates an inconsistent conceptualisation of the theory between articles. Further, after publishing this 2018 paper claiming that DFD predicts changes in showup performance, Colloff and Wixted (2019) argued that show ups do not allow for the DFD process to play out: The witness cannot determine which features are shared and which are important for making the identification. There is no comparison across faces and thus the witness may rely too heavily on non-diagnostic features. It is concerning that DFD has been presented in such different ways between articles. Colloff et al.'s (2018) results could not provide support for filler-siphoning, but also could not provide support for DFD, despite the authors' conclusions.

Colloff and Wixted (2019) introduced a novel lineup procedure to ensure that differential filler siphoning could not be presented as an explanation, while DFD remained a viable theory for explanation. In the "simultaneous show up" condition, participants viewed a six-person lineup made up from similar looking fillers. This lineup became a show up when the suspect was highlighted within the lineup, with a red banner around the image. This allowed for direct comparison of the only suspect to the other lineup members and was theorised, according to DFD, to enhance discrimination of guilty from innocent suspects. This procedure only required a yes/no response to the suspect. This procedure and the standard simultaneous lineup enhanced discriminability compared to presenting the suspect in a standard show up.

Interestingly, Collof & Wixted (2019) did not incorporate a sequential lineup for direct comparison to their procedure. Nonetheless, these are the first findings that provide direct support for DFD theory (though not as a mechanism for differences between sequential and simultaneous presentation) in the eyewitness context.

Similarity Effects on Facial Recognition

To provide diagnostic outcomes, a lineup needs to be fair to the suspect to prevent them from standing out. However, very high levels of suspect-filler similarity can produce a very difficult task for the witness (Fitzgerald, Oriet & Price, 2015; Smith, et al., 2018). In lineup construction, we need to consider how physically similar fillers should be in relation to the suspect (Clark, 2003; Malpass, Tredoux & McQuiston-Surrett, 2007). The fillers need to be viable options based on physical similarity, however Luus & Wells (1991) cautioned that if fillers are too similar to the suspect, the recognition task will be too demanding for the witness and thus reduce accuracy (Tredoux, 2002). Brewer & Wells (2006) proposed that in target present lineups we should expect to see increased correct identifications when low similarity fillers are used, compared to high similarity fillers. In target absent lineups, the use of low or high similarity fillers will increase the risk of the plausible innocent suspect being identified, due the relative familiarity mechanism. Tredoux's (2002) findings were consistent with the notion that increased similarity reduces accuracy. The low-similarity lineups led to greater accuracy in terms of correct identifications and rejections, in comparison to moderate- and high-similarity foils.

Tulving (1981) investigated the standard similarity effect with complex images of landscapes, buildings, and groups of animals as the stimuli. Memory was tested with a two alternative forced choice task, in which the participants were asked to identify which of two images they had seen previously. The three conditions of

this study concerned the similarity between target and distracter, at test. In the first condition, the lure and distractor were similar, meaning that one half of the original image was the target and the other half of the image was shown as the distractor. In the second, the distractor was dissimilar to the target, but was similar to another previously seen image. In the third, the distractor was dissimilar to the target and other previously seen images. The results showed the standard similarity effect, whereby recognition accuracy was higher when the distractors were dissimilar to targets and other items presented at study, compared to when they were similar. However, participants were more accurate at discriminating between distractors and targets when they were perceptually similar (in condition one) compared to when the distractor was dissimilar to the target but similar to other previously seen images (condition two). Overall, Tulving's (1981) experiments showed that accuracy in this type of recognition task was higher when the distractors were perceptually very similar to the target, rather than when a lower level of target-distractor similarity or a dissimilar distractor was used. Tulving proposed that when two similar items are presented and share the same memory trace, the person is able to disregard shared features and focus on the features that distinguish a target from a distractor by matching these features to the memory trace. Tulving's explanation for these findings actually aligns quite neatly with DFD theory and the findings provide theoretical support for the current study to investigate DFD with a manipulation of target-lure similarity.

Horry and Brewer (2016) investigated how target-lure similarity influences confidence judgements in five experiments using facial recognition tasks. In the two alternative forced choice task, participants had to choose a face (i.e., all target present trials). In this condition the standard similarity effect was observed, with

discriminability improving as target-lure similarity decreased. In their third experiment, a compound decision task was used, in which participants could choose one of two faces, or reject the array. In this case, the standard similarity effect was observed in target present trials, whereby increased similarity made the task more difficult and reduced discrimination. However, in target absent trials, target-lure similarity did not influence willingness to choose (response bias) or ability to distinguish target absent from target present trials.

Fitzgerald et al.'s (2013) meta-analysis explored how differing levels of suspect-filler similarity affected identification decisions. Identifications of both innocent and guilty suspects were most common in low similarity lineups. Further, high similarity lineups (compared to moderate) protected against false identifications of innocent fillers, without interfering with correct identifications of the culprit. This indicates that the use of high similarity fillers could be beneficial, rather than detrimental. Finally, filler identifications occurred more frequently from moderate and high similarity lineups, independent of whether the culprit was present or absent. This finding aligns with those discussed in regard to Colloff et al. (2016). Fitzgerald et al. (2015) manipulated suspect-filler similarity using morphing software to create moderately high and very high similarity lineups. The very high similarity lineup had low rates of correct identifications and increased rates of filler identifications, indicating that morphing to that level had made the task too difficult. To summarise, there is likely there is no ideal level of similarity for fillers to be in comparison to the suspect, as filler similarity will likely interact with suspect plausibility (and guilt) in effects on performance. Generally, increased similarity makes recognition memory and perceptual discrimination tasks more difficult, but DFD suggests simultaneous presentation might attenuate these effects (Wixted & Mickes, 2014).

Reynolds et al. (2019) tested the use of the 'don't know' response, that allows people to opt out from making an identification, in a study that incorporated a manipulation of lure similarity. In this study, participants studied a series of faces, followed by a retention interval and test phase. Participants were randomly allocated to one of two response type conditions, either making a yes/no (Y/N) or two-alternative forced choice (2AFC) recognition decision for each test face, with the option of responding with 'don't know'. In the Y/N condition, one photo (either a studied/target face or non-studied lure) was presented at a time and participants were asked whether they recognised the face. In the 2AFC condition, participants were presented with two faces per trial, one of which was always a target face. Participants were asked which face they recognised. Difficulty was manipulated through high and low levels of similarity between the targets and lures. This manipulation produced standard effects on accuracy for Y/N decision (i.e., lower accuracy for high-similarity lures) but did not affect performance in the 2AFC condition. Participants could distinguish the targets from high and low similarity lures equally well when the faces were presented simultaneously. The authors noted these findings as consistent with DFD theory, and the idea that simultaneous presentation allows for shared features of the stimuli to be discounted, allowing individuals to notice features that are diagnostic of guilt (Wixted & Mickes, 2014).

However, this is not the only explanation. As the 2AFC condition only used target present trials, the null effects of similarity on performance could be accounted for by a solely relative familiarity decision strategy. In this task, participants simply needed to decide which of the two options was *most* familiar. That is, there was no "absolute" component required for this decision. In this case, the familiarity of the target in each trial was seemingly strong enough that participants were able to choose correctly most of the time.

We know that relative decisions often produce accurate identifications when the culprit is there to be recognised (i.e., in target present lineups; Dobolyi & Dodson, 2013; Steblay et al., 2001; Steblay et al., 2011), but we also know that this approach increases errors when the target is absent. All recognition decisions rely on a sense of familiarity. Simultaneous presentation facilitates a comparison of response options for *relative* familiarity. Indeed, DFD hypothesis could be framed as improving the diagnostic value of the sense of familiarity participants experience, by helping them to base this sense on diagnostic features. In this way, relative judgements might facilitate DFD in target present lineups. Alternatively, decisions might simply reflect the relative familiarity mechanism. In a 2AFC task, decisions based solely on this relative familiarity mechanism can provide good discrimination, because the target is always there to be recognised. When participants must entertain the possibility that the target is absent, the use of this relative familiarity mechanism will not be effective as the most familiar option will not be the target. If DFD theory provides the best explanation for Reynolds' et al. findings and DFD actually improves discrimination in a way that is meaningful in an applied context (i.e., lineups), then it should improve participants' ability to tell which face, if any, they have seen before, and not just improve their ability to correctly choose the face from target present lineups. Demonstrating null effects of similarity in a 2AFC task is consistent with the DFD hypothesis, but this does not indicate that simultaneous lineup presentation will improve discrimination in a task that requires individuals to entertain the possibility that *none of the presented options* is the target. Determining whether these findings support DFD theory or whether they are solely due to the appropriateness of relying on relative familiarity formed the main aim and rationale for the current study.

The Current Study

The current research replicated Reynolds et al.'s (2019) Y/N and 2AFC conditions (with the exception that we did not include a "don't know" response option) and introduced a new condition central to testing the diagnostic feature-detection hypothesis. Specifically, we included a two-alternative open choice (2AOC) condition, also known as a compound decision. A compound decision is more complex than a 2AFC task (Duncan, 2006; Horry & Brewer, 2016). Whereas a 2AFC task requires a participant to determine which test stimulus is the target, a 2AOC task requires the participant determine which, *if any*, test stimulus is the target. . In this condition half of the trials were target absent, all trials included a "not present" response option, and participants were informed that the target may or may not be present in each array (Duncan, 2006). In the 2AOC, making a simple relative familiarity judgment is no longer a viable strategy as, in some cases, the most familiar item will nonetheless be a lure. Incorporating this new condition allows us to investigate whether the similarity effect will re-emerge (e.g., if the previous null effect of similarity was due to the appropriateness of the simple relative familiarity judgement strategy) or whether simultaneous presentation still attenuates similarity effects (i.e., as according to DFD).

SDT is used to separate discriminability from response bias to understand overall performance. We did this by calculating indices for these two parameters: *d'* and *c*, respectively (Colloff et al., 2018; Macmillan & Creelman, 2004; Palmer & Brewer, 2012). The process of calculating these parameters is straightforward in simple yes/no recognition decisions in which there are only four possible responses (hits, misses, false alarms and correct rejections). However, this becomes more complex when different types of recognition decisions required; specifically, when there are multiple options available and a filler could be identified. This is the case

for the 2AFC and 2AOC conditions in the current study. One issue is how responses should be coded when a filler is identified from a target present lineup: as a false alarm (filler identification) or a miss (culprit missed). Filler identifications might not be forensically relevant as they do not lead to miscarriages of justice, nevertheless they need to be taken into consideration when describing recognition performance, as they are psychologically relevant (i.e., they represent a memory failure; Duncan, 2006; Palmer & Brewer, 2012).

Further, 2AOC is a more complex, compound decision task, comprised of detection and identification components (Horry & Brewer, 2016; Palmer, Brewer, & Weber, 2010). Thus, performance indices (*d'* and *c*) for the compound decisions made by participants in the 2AOC condition were calculated using a signal detection theory compound decision model (SDT-CD; Duncan, 2006; Palmer, Brewer, & Weber, 2010). SDT-CD is capable of taking into consideration filler identifications from target present and absent lineups, and thus produces a better description of identification performance in compound decisions tasks than standard signal detection theory models which allow for only four types of decisions (i.e., hits, false alarms, correct rejections, and misses).

Hypotheses

Based on Reynolds et al.'s (2019) findings and the fact that our stimuli were nearly identical and procedures were very similar, we expected to replicate the standard similarity effect for the Y/N task, whereby participants would be less accurate when responding to high similarity compared to low similarity lures. In terms of similarity effects in the 2AFC and 2AOC conditions, we had competing hypotheses. Both of these hypotheses rely on a relative familiarity mechanism. If a relative familiarity mechanism allows for the discounting of non-diagnostic features

and the detection of diagnostic features, we would expect, consistent with the findings from Reynolds et al. and DFD, smaller or absent effects of similarity on discrimination in both the 2AFC and 2AOC conditions, compared to the YN condition (Wixted & Mickes, 2014). On the other hand, if a relative familiarity mechanism simply makes it easier for the participant to determine which of two response options is most familiar (without necessarily making it easier to determine that neither option was previously seen) we would expect smaller effects of difficulty in only the 2AFC, not the 2AOC and YN conditions (see *Figure 2* for expected pattern of results).

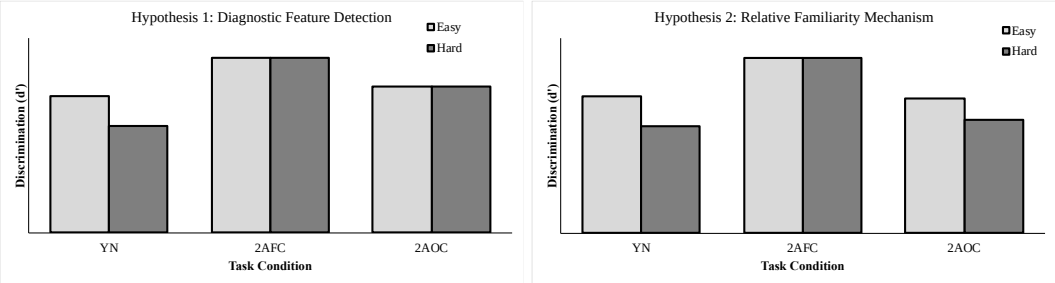

Figure 2. Expected patterns of results, based on DFD and simple relative familiarity mechanisms.

Method

Design

We used a 2 (target-lure similarity: high or low) x 3 (task condition: Y/N, 2AFC, 2AOC) mixed factorial design, with target-lure similarity as the repeated measures variable. The 2AFC condition contained only target present trials, whereas the Y/N and 2AOC conditions included both target present and absent trials. The dependent variable was recognition performance, measured by discrimination (*d'*), which measures the individual's ability to discriminate between faces seen before and faces not previously seen.

Participants

In accordance with Reynolds et al. (2019), we recruited 60 participants (42 women) aged 18 to 45 years (M=25.8, SD=5.71). Each participant provided 288 data points, supporting the precision of our within-subjects measures (Charness, Gneezy, & Kuhn, 2012). Participants were over the age of 18 with normal or corrected-to-normal vision. Most participants were students of the University of Tasmania, invited through advertisements placed around the Sandy Bay campus. First year psychology students also signed up using SONA and received one hour of research credit. Other participants were reimbursed with a $30 Coles/Myer or Gift Pay voucher. Data collection took place in a Tasmanian Cognition Laboratory testing room containing multiple computers.

The research was approved by the Human Research Ethics Committee, University of Tasmania (approval number: H0018178).

Materials

The experimental software was created with Java (JDK7) and the libGDX graphics framework. Participants completed the study on desktop computers equipped with 3.30 Ghz Intel i5-6600 processors, 16 GB RAM, and a Windows 7 enterprise operating system configured to minimise internal task-switching. The program was displayed on 24-inch monitors.

Stimuli

Our stimuli were sourced from Reynolds et al. (2019). Originally, the stimuli came from our laboratories database and two additional face databases (Burton, White & McNeil, 2010; Thomaz & Giraldi, 2010). Face stimuli were predominately Caucasian, with a small amount of Asian, Middle Eastern and South American descent. The stimuli mainly included the head and shoulders in the image, with a

small proportion including only the head and neck. Test faces were presented on a grey and white patterned background, which differed from study phases to encourage face recognition, rather than picture recognition (Reynolds et al., 2019). The background image was 15cm x 12cm.

To manipulate difficulty across the conditions, two levels of target-lure similarity were created through a process of morphing two faces, using the face morphing software FantaMorph 5 (Abrosoft, 2016). The new faces were generated by combining differing ratios of two faces. The faces were paired based on hair colour and style, and face shape. The high-similarity lure was made up from 30% of the target face and 70% of a new face (see *Figure 3*). The low-similarity lure was 10% target face morphed with 90% of a new face (see *Figure 4*). Feedback from a previous study established these two levels as the hard and easy to discriminate conditions (Reynolds et al., 2019).

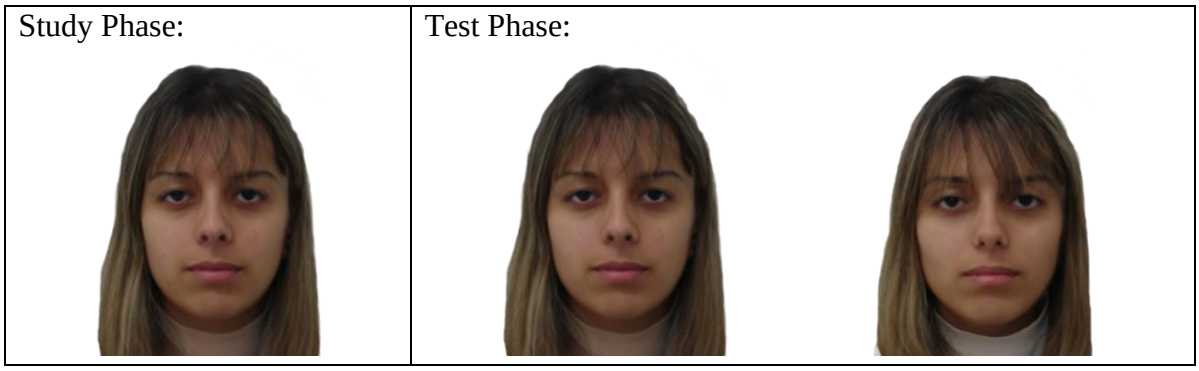

Figure 3: Example hard (high similarity) trial.

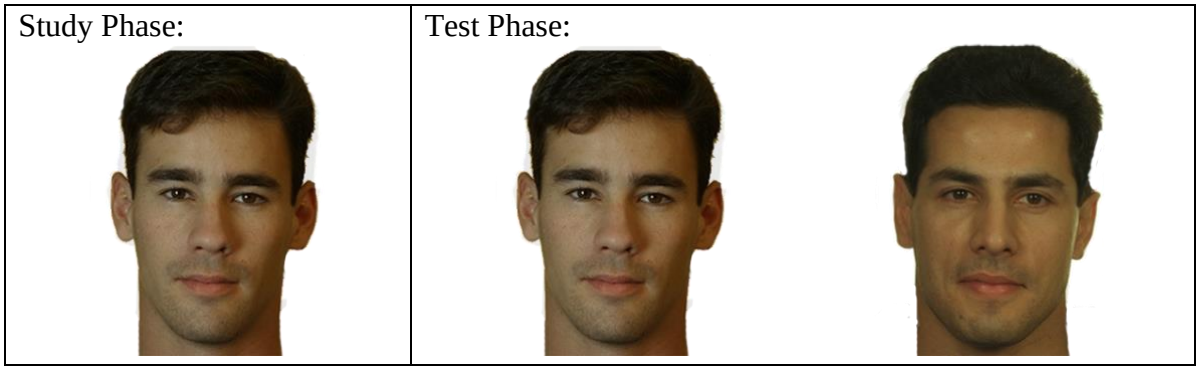

Figure 4: Example easy (low similarity) trial.

Procedure

Participants read the information sheet and signed the consent form. Minimal verbal instructions were required, as most instructions were displayed on-screen. On-screen instructions included an image of the correct placements for fingers on the keyboard that showed participants how to place both thumbs on the spacebar and their index or middle fingers on the "z" and "/?" keys.

Participants were randomly allocated to one of the three task conditions. All conditions included a practice block of one study phase and one test phase. Following this, participants in the Y/N and 2AFC conditions completed 12 blocks of trials, each block consisted of 24 study items and 24 test trials. Participants in the 2AOC completed six blocks of trials, with 24 study items and 48 test trials per block. The 2AOC condition had half as many blocks, because the test phase involved twice as many test items compared to Y/N and 2AFC. It was important to keep the number of study items constant across conditions, to ensure we did not interfere with cognitive load at study between conditions. If the number of study items was increased in the 2AOC condition, it would have added a confound (i.e., difficulty) to the manipulation of condition.

At study, 28 faces were presented, with four of those being buffer items (two shown at the beginning and end of the phase) to reduce primacy and recency effects (Ebbinghaus, 1885). Responses to these items were not included in analyses. Of the 24 testable faces, half were drawn from the low-similarity pairs and half from the high-similarity pairs. A central fixation point was shown for 0.5 seconds prior to the presentation of the first face. Each face was then presented in the centre of the screen for 1s. A blank screen was shown for 0.5s as an interstimulus interval (ISI). The test

phase began 1s after the final face was shown at study. The first test trial was a buffer item, and thus, excluded from analyses.

In the Y/N condition, at test participants viewed one face at a time and had to decide if they recognised the face (yes/no). In the Y/N condition, 12 faces were previously seen targets, and 12 faces were new: six were low-similarity lures and six were high-similarity lures. In the 2AFC condition, each trial presented two faces side by side and participants picked which one they had seen before. Every trial contained a target. In the 2AFC condition, all 24 study items were tested, 12 were presented with a high-similarity lure, and 12 presented with a low-similarity lure. In the new, 2AOC condition, at test participants viewed two faces at a time and were asked to make a two alternative open choice decision and choose 'A', 'B' or 'not present'. All 24 study items were tested, in addition to 24 new faces for the target absent trials. The 48 test items were presented in the same method as the 2AFC, with their corresponding lures (half high-similarity, half low-similarity).

During initial instructions, all participants were informed that a time limit of 2.5s per face would be enforced and were provided with an example of this with a blank screen shown for 2.5s. Each test phase began with a central fixation point displayed for 1s. Test items were displayed for a maximum of 2.5s, after which the item would timeout and the participant was informed of this with the message: 'Ran out of time!'. If participants responded correctly, they received a message informing them they were correct and had received points ('+100 points' in green). If they responded incorrectly, they received a 'wrong' message and lost points ('-100 points' in red). Points were intended to act as a motivator for participants (Reynolds et al., 2019). The number of points displayed quickly moved to the top right of the screen where the accumulated score was displayed during test phase. The overall score

could not drop below zero, to circumvent motivation loss. Points accumulated over blocks of trials and the total score was displayed at the end of each block and at completion of the study. After completing the study, participants were debriefed and thanked.

Results

Analysis strategy

Our interest was in assessing whether, as predicted by a DFD hypothesis, simultaneous presentation attenuates the effects of increased similarity on task performance. Although answering this question involves only a simple comparison of task performance across three conditions, there are a number of ways to measure task performance. The simplest would be to look at the effects of similarity on mean accuracy in the three conditions. However, accuracy conflates two key mechanisms underlying performance: discrimination (*d'*) and bias (*c*). Our research adopted a SDT-based approach to investigate the effects of task type on recognition performance, which teases apart these two mechanisms (Macmillan & Creelman, 2004). We were primarily interested in difficulty-related changes in *d'* within each condition. Larger *d'* values imply better performance (Macmillan & Creelman, 2004). To assess difficulty effects on *d'*, we calculated *d'* for each participant for overall performance, and for performance on easy and hard trials. We then averaged *d'* within each task type condition, for overall performance and performance in easy and hard trials. For participants in Y/N and 2AFC conditions, simple *d'* was calculated. As the decision task in the 2AOC was more complex, compound decision *d'* was calculated for participants in this condition using the SDT-CD model (Duncan, 2006; Palmer & Brewer, 2012).

This model proposes that compound decisions are comprised of two components: detection and identification. In this model, detection performance is modelled as 1-of-*m* detection, where '1' indicates the culprit may or may not be present amongst *m* number of fillers. This part of the compound decision requires discrimination between a target present and target absent lineup, but not the identification of the culprit (Duncan, 2006). Identification is modelled as an *m*-alternative forced choice decision task, in which the culprit must be identified from the array of *m* fillers. SDT-CD calculates estimates of response probabilities and compares these to the observed responses probabilities. Through comparison of the response probabilities, we can establish the most suitable combination of discrimination and response bias that describes the performance in the observed data (Palmer & Brewer, 2012). See Tables 3, 4 and 5 for individual participants SDT-CD calculations for the overall, easy and hard trials, respectively (in Appendix A). Simple *d'* and compound decision *d'* cannot be compared to each other, due to the differences in the decision task on which they are based (Duncan, 2006; Palmer & Brewer, 2012). As we were investigating how the effect of difficulty would vary according to condition, it was not necessary to compare *d'* across the conditions, only to test for effects on *d' within* conditions.

In our study, participants had 2.5 seconds to make each recognition decision, after which the trial would time out. We lost 2.36% of trials across all conditions to these time outs. The most trials were lost in the 2AOC (3.47%), followed by 2AFC (2.35%) and Y/N (1.28%).

A mixed ANOVA was used to answer our key question: did the effect of difficulty (similarity level) vary according to task type. Thus, the key effect of interest was the Difficulty × Task Type interaction, not the main effects (though, for

completeness, we present the for main effects of difficulty and task type in Table 1). We used partial eta squared as our measure of effect size for our ANOVA, which is interpreted based on the benchmarks for small, medium and large set by Cohen (1988) as .01, .06, and .14, respectively. For the follow-up simple effects analysis, we used Hedge's *g*, the bias-corrected version of Cohen's *d*, as our measure of effect size; Cut offs for small, medium and large are 0.2, 0.5, and 0.8, respectively (Ellis, 2010; Hedges, 1981).

Testing the DFD: Effects of difficulty and task type on discrimination

As a reminder, according to the DFD hypothesis we would expect to see smaller or absent effect of similarity on discrimination in both the 2AFC and 2AOC conditions. Alternatively, according to the relative familiarity mechanism, we would expect to see smaller or absent effect of similarity on discrimination in the 2AFC only. Both of these mechanisms would manifest as a significant Difficulty x Task Type interaction, and we would expect the differences to emerge in follow-up simple effects analysis. Specifically, DFD would predict no differences in discrimination between easy and hard trials in both the 2AFC and 2AOC, whereas, relative familiarity mechanism would only predict this in the 2AFC.

A 3(Task Type: YN, 2AFC, 2AOC) × 2(Difficulty: easy, hard) mixed ANOVA with Task Type as the between-subjects variable, returned a significant Task Type × Difficulty interaction (see Table 1 for inferential statistics and *Figure 5* for descriptive statistics. This interaction was followed up with three paired samples t-tests (see Table 2 in Appendix A for descriptive statistics), which revealed that for the Y/N condition, there was no statistically significant difference in discrimination between the easy and hard trials, $t(19) = 1.285$, $p = .214$, $g = 0.20$. In contract, the difficulty manipulation exerted large, significant effects on discrimination in the

2AFC condition, $t(19) = 8.159$, $p < .001$, $g = 0.87$, and the 2AOC condition, $t(19) = 7.117$, $p < .001$, $g = 0.79$. Thus, counter to both the DFD-based and familiarity mechanism-based hypotheses, similarity exerted large effects on discrimination in the simultaneous presentation conditions.

Table 1

Results of the 3(Task Type) x 2(Difficulty) mixed ANOVA

	df	*F*	*p*	η_{p2}
Difficulty (within-subjects)	1	75.718	<.001	.571
Task Type (between-subjects)	2	3.445	0.039	.108
Interaction	2	9.196	<.001	.244

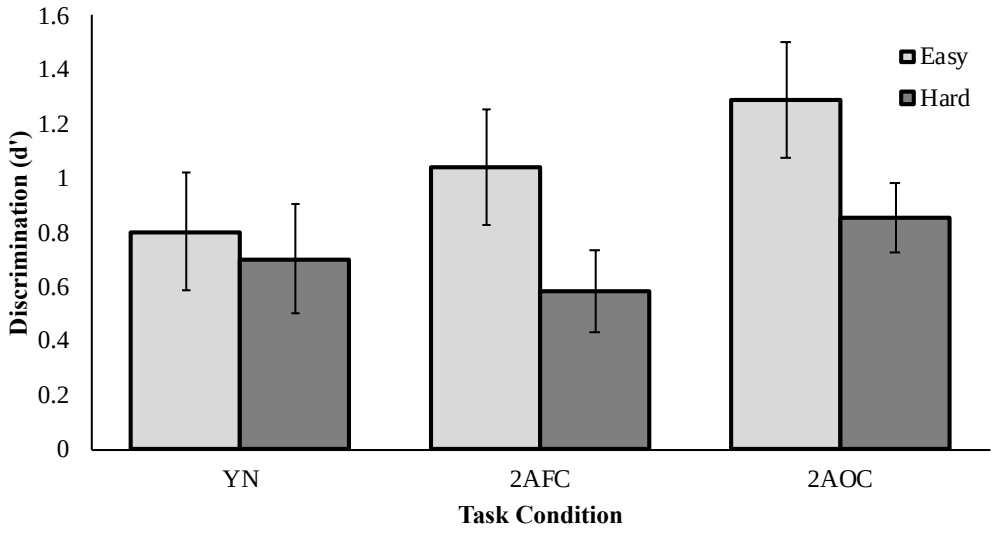

Figure 5. Discrimination in easy (low similarity) and hard (high similarity) trials for each task type. Errors bars represent 95% CIs.

As indicated above, our question would be answered by comparing the effects of difficulty on discrimination across task conditions, rather than focusing on absolute levels of discrimination within task conditions. However, the non-significant effect of difficulty in the YN condition was surprising. Thus, to test whether this non-significant difference was associated with a genuine null effect we conducted a Bayesian paired samples t-test using a default Cauchy prior of .707. The BF_{01} was 2.1, indicating only anecdotal evidence for the null hypothesis. Thus, although this difference was non-significant, the Bayesian analysis did not provide compelling evidence for the null. Nonetheless, the effect of difficulty was smaller in the Y/N than in the 2AFC and 2AOC conditions.

Exploratory analysis: Effects on response bias

We are not interested in bias in the 2AFC condition, because in this type of task bias only indicates tendency to choose the left or right option. We were interested in how difficulty may have affected response bias differently in the two conditions where participants needed to decide if a previously seen face is present or absent. To investigate this, two paired samples t-tests were utilised. The difficulty manipulation exerted large effects on response bias in the Y/N condition, with a significant difference in bias between easy (M=.01, SD=.19) and hard (M=-.17, SD=.22) trials, $t(19) = 5.26$, $p < .001$, $g = .84$. This indicates participants were more lenient in responding to hard trials. In the 2AOC condition, there was no significant difference in response bias between the easy (M=.13, SD=.31) and hard (M=.16, SD=.28) trials, $t(19) = -.902$, $p = .378$, $g = .11$.

Discussion

Understanding the effects of lineup presentation method on identification performance has long been of interest to researchers aiming to reduce eyewitness

identification errors (Cutler & Penrod, 1995; Sauer et al., 2019). Recently, counter to the traditional view that sequential presentation improves performance, researchers have argued that simultaneous presentation may improve discrimination: helping witnesses discriminate between guilty and innocent lineup members. Diagnostic feature-detection (DFD) theory proposes that simultaneous presentation facilitates the discounting of shared features and detection of diagnostic features (Mickes et al., 2012; Wixted & Mickes, 2014). Recent findings suggest that simultaneous presentation might also attenuate the effects of increased target-lure similarity on task performance in a 2AFC recognition task, consistent with DFD (Reynolds et al., 2019). However, these findings could also be explained by a simple relative familiarity mechanism, whereby relative comparison allows people to identify the most familiar member in the array. This would improve discrimination in a 2AFC task, where the target is always present, but not necessarily in a task where the target might be absent (as in a lineup). Thus, we directly tested the DFD hypothesis by comparing similarity effects on discrimination in Y/N, 2AFC and 2AOC recognition tasks.

Tests of the DFD and Relative Familiarity Mechanism

In accordance with DFD, compared to the Y/N condition, we expected to observe smaller effects of similarity on discrimination in both the 2AFC and 2AOC, where simultaneous presentation facilitates relative judgments and the detection of diagnostic features and discounting of non-diagnostic features. Based on a simple relative familiarity mechanism, we would expect to find smaller effects of similarity on discrimination in the 2AFC (consistent with Reynolds et al., 2019, and work in the identification domain; Dobolyi & Dodson, 2013; Steblay et al., 2001; Steblay, 2011), but not in the 2AOC, where this strategy will not help participants detect the

target's absence. We found no evidence of smaller effects of similarity in either of the simultaneous presentation conditions (2AFC; 2AOC). In fact, there were large, significant effects of similarity on discrimination in both conditions. Moreover, these effects were larger than in the Y/N condition. Discrimination may have been better in the simultaneous presentation conditions (but, as previously explained, we cannot directly compare *d'* across the conditions). The important point here is that simultaneous presentation did not attenuate the deleterious effects of increased similarity on *d'* (compared to the Y/N condition). These findings provide no evidence for the DFD or simple relative familiarity mechanisms. Previous researchers have made strong claims about the diagnostic benefits of simultaneous presentation and proposed DFD as the mechanism at play, without conducting any direct tests (Wixted & Mickes, 2014; Colloff et al., 2016; Colloff & Wixted, 2019). We provided the first direct test of the DFD hypothesis, and found no support for this mechanism.

Based on Reynolds' et al. (2019) findings, we expected the Y/N condition to show the standard similarity effect: reduced discrimination between seen and previously unseen faces in the hard (cf. easy) trials. Unexpectedly, we did not replicate Reynolds' et al. findings: we found no evidence for an effect of similarity on discrimination in this condition. While not the key focus of our study, this finding was unanticipated and warrants consideration in terms of its implications. We are cautious to give too much weight to this finding given it flies in the face of established memory effects (Tulving, 1981), and given our Bayesian analysis did not provide compelling evidence for the null. However, the fact remains that, even if the current data underestimated the similarity effect in the YN condition, there were still large effects of similarity in the other conditions. Thus, it would take a large

underestimation to overturn our conclusion that the effect of similarity was larger in the simultaneous presentation conditions than in the Y/N, and that this pattern of results provides evidence against the DFD mechanism.

In trying to understand why we did not replicate Reynolds' et al. (2019) findings, there are a number of points to consider First, Reynolds et al.'s finding of no effect of similarity in the 2AFC was surprising in and of itself. So, in a sense, the failure to replicate this aspect of their findings is not surprising. However, our intent was not necessarily to replicate this finding. We were investigating whether we could replicate a smaller effect of similarity in the 2AFC task, compared to the Y/N task. Our failure to replicate this aspect of their findings requires some consideration. One potential explanation could be differences in design. The task conditions in the current study were not exact replications of those used in Reynolds et al.. First, Reynolds et al. included other manipulations, such as the speed vs. accuracy emphasis, which could have influenced participant responding. However, there is no theoretical reason for this speed accuracy manipulation to interact with the DFD mechanism. The 2AFC in the current study was a strict two alternative forced choice task, in which participants had to choose one of the two faces.

Second, Reynolds et al. (2019) included a 'don't know' (DK) response option, and we did not. Including a DK response option might have improved performance. According to Weber and Perfect (2013), eyewitnesses are unlikely to use a DK response, unless it has been made clear that it is acceptable to do so. Without this knowledge, eyewitnesses make decisions they are uncertain of. Including a 'don't know' response has been found to improve the accuracy of identification decision (Weber & Perfect, 2013). The presence of a DK response allows the individual to correctly recognise that they cannot make a useful

judgement, leading to a better quality decision. Perhaps in Reynolds et al., having the option to opt out of the decision lead to more conservative responding, whereby participants would only choose a face if they were certain they had seen it before. However, critically and in regard to DFD, there is nothing in the literature to suggest a DK *option is a necessary precondition for the DFD mechanism*. Simultaneous presentation is the key to DFD. According to DFD, if the faces are presented in an array, side-by-side, this is will allow shared features to be discounted, and diagnostic features to be recognised (Wixted & Mickes, 2014). This logic does not require a DK option.

Notwithstanding the differences between our design and Reynolds et al., there were a number of aspects that were kept consistent between the studies: we utilised the same stimuli and points system to keep participants motivated as Reynolds et al. The Y/N and 2AFC tasks had been pilot tested for difficulty (although the 2AOC was not) and, in both studies, each participant completed hundreds of trials, indicating the differences in our results were not due to an issue of power (Charness et al., 2012). Thus, there seems little reason to believe that design choices for the present study are responsible for our inability to replicate their findings.

Implications for Lineup Construction

Finding an effect of similarity in our simultaneous presentation conditions aligns with previous research that indicates increased similarity between the suspect and lineup members increases the difficulty of the task, thereby reducing accuracy (Brewer & Wells, 2006; Horry & Brewer, 2016; Luus & Wells, 1991; Smith, et al., 2018; Tredoux, 2002). In lineup construction, we need to consider how similar lineup member should be to each other to make the lineup fair (Colloff et al., 2016).

Although major implications cannot be drawn from the current data, our results do indicate that simultaneous presentation does not necessarily negate the deleterious effects of similarity in recognition decisions, as predicted by DFD.

Limitations and Future Directions

We utilised a stimulus set that was created and piloted for difficulty in 2016 (Reynolds et al., 2019). Participants were found to be capable of discriminating between targets and lures in the 2AFC condition. However, the stimuli had not been piloted in the 2AOC condition for the current study and as this condition involves a more complex decision task (i.e., compound decision), the task may have become too difficult. According to Fitzgerald et al. (2015), a pair of faces morphed at 30%/70% is too difficult, suggesting our high similarity lures were too difficult. However, discrimination was not at floor (~.8 in the 2AOC hard trials) and participants could do the Y/N task, suggesting we were not seeing catastrophic failures of memory. This suggests that the absence of expected effects was not due to problems with the stimuli used.

Perhaps the absence of expected effects reflects the chosen measure of discrimination to answer our key questions. This is only one measure of performance and other techniques could be employed - overall accuracy, diagnosticity, and ROC analysis – and different approaches produce different conclusions. For instance, Mickes et al. (2012) and Wixted and Mickes (2014) ROC analyses showed simultaneous lineups had a higher level of discrimination compared to sequential, whereas Palmer & Brewer (2012) use of discrimination and bias showed that sequential lineups affect bias, but neither sequential nor simultaneous lineups were better in terms of discrimination. As previously mentioned, Mickes et al. have argued that by condensing discrimination down to a single point, diagnosticity ratios can

mask meaningful effects. The same may be true when researchers consider discrimination and bias based on the point estimates *d'* and *c*. Perhaps this accounts for the different patterns of results reported by Palmer & Brewer (2012) using *d'* and *c*, and Mickes et al. (2012) using ROC analysis. Our analytical approach – focusing heavily on *d'* as our index of discrimination – might similarly mask patterns that would be evident based on ROC analysis. This may be indicated by the effect of bias in the Y/N task, but no effect of discrimination, and the opposite in the 2AOC task, in which there were effects on discrimination, but not bias. However, conducting ROC analyses was not an option in the current research as we did not collect confidence ratings. However, given the differential effects on discrimination and bias for the YN and 2AOC conditions, re-running the study and collecting confidence ratings to allow for ROC analysis (and reducing the chance that relying on point estimates of d' and c obscured effects) might be beneficial.

From a theoretical perspective, the DFD hypothesis requires development as its predictions are ambiguous for target absent lineups. DFD does not outline how a witness recognises that the culprit is not present. We need a better understanding of what DFD would predict in a target absent lineup. Currently, it is unclear whether DFD primarily expects improvements in discrimination in target present lineups only, or whether the benefits would extend to rejecting target absent lineups. The expected effects need to be teased apart to form a better understanding of DFD theory.

In spite of the current failure to find support for DFD, we note this was the first direct test of DFD, and thus further research is warranted. Faces are inherently complex, especially in comparison to stimuli used in basic recognition and discrimination research. For this reason, it may be beneficial to start by investigating

recognition for simpler stimuli to build our understanding of the extent to which DFD generalises from perceptual discrimination to recognition memory tasks. As there is evidence for DFD in perceptual discrimination tasks, we might start with recognition of basic stimuli, such as arrangements of basic shapes or geometric patterns. The arrangement/shapes could be altered to create two subjectively discrete levels of similarity (e.g., changes in colour, size of shapes). Such manipulations would be easily achievable and amenable to tight experimental control. Through this, an understanding of the sorts of stimuli and tasks for which DFD is likely to appear can be developed. This will support a better understanding of DFD and its applicability to complex recognition tasks (e.g., lineups of faces).

Conclusions

Eyewitness researchers aim to identify procedures that maximise identification accuracy and, critically, develop a theoretical understanding of the mechanisms that drive performance. Despite the claims of researchers in the field, our results – representing the first direct test of the DFD as an explanation for improved discrimination under simultaneous (cf. sequential) presentation conditions – do not support a DFD mechanism.

Understanding the mechanisms underlying identification performance is critical because otherwise ideas that seem clever (if a theoretical account is true) can actually be dangerous (if the theoretical account is false). For example, consider Colloff and Wixted's (2019) novel procedure. As a reminder, the novel procedure, the 'simultaneous showup', is initially presented as a standard simultaneous lineup. The innovative feature of this procedure is the showup aspect, which occurs when the suspect is highlighted within the array. This procedure informs the witness which person they need to focus on: thereby preventing fillers identification, while also

highlighting shared, non-diagnostic features. Thus, DFD would expect the witness to be capable of comparing the suspect to the other lineup members and identifying diagnostic features to determine whether the suspect is the culprit. If DFD is viable, this would be a clever way of utilising the DFD mechanism and could be effective in promoting accurate identifications. However, if not, the researchers have essentially created a biased and suggestible lineup.

In sum, our study provides no support for DFD theory, as simultaneous presentation of highly similar faces did not improve discrimination, nor did it negate the similarity effect. It also provided no support for the simple relative familiarity mechanism. Although previous work suggest simultaneous presentation can improve discrimination (Gronlund et al., 2009; Gronlund et al., 2012; Mickes et al., 2012), we found no evidence it would protect discrimination from the effects of highly similar lures. Due to the absence of direct empirical support, DFD theory cannot currently explain the improved discrimination associated with simultaneous presentation, despite previous authors' claims (e.g., Colloff & Wixted, 2019; Wixted & Mickes, 2014). Further research is required to investigate the theoretical value of DFD, and generally to understand the mechanisms through which simultaneous presentation leads to improved discrimination.

References

Abrosoft (2016). Abrosoft FantaMorph (5). Abrosoft. License obtained from http://www.fantamorph.com/index.html

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30*(1), 31-50. doi: 10.1007/s10979-006-9002-7

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*(1), 77-96. doi: 10.1348/135532509X414765

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*(1), 11-30. doi: 10.1037/1076-898X.12.1.11

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42(*1), 286-291. doi:10.3758/BRM.42.1.286

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1-8. doi:10.1016/j.jebo.2011.08.009

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *17*(6), 629-654. doi: 10.1002/acp.891

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). HilIsdale, NJ: Lawrence Erlbaum Associates.

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27*(9), 1227-1239. doi: 10.1177/0956797616655789

Colloff, M. F., Wade, K. A., Strange, D., & Wixted, J. T. (2018). Filler-Siphoning Theory Does Not Predict the Effect of Lineup Fairness on the Ability to Discriminate Innocent From Guilty Suspects: Reply to Smith, Wells, Smalarz, and Lampinen (2018). *Psychological Science, 29*(9), 1552-1557. doi: 10.1177/09567718786459

Colloff, M. F., & Wixted, J. T. (2019). Why are Lineups Better than Showups? A Test of the Filler Siphoning and Enhanced Discriminability Accounts. *Journal of Experimental Psychology: Applied*. Advance online publication. doi: 10.1037/xap0000218.

Cutler, B. L., & Penrod, S. D. (1995*). Mistaken identification: The eyewitness, psychology and the law.* Cambridge: Cambridge University Press. Retrieved from https://www.ncjrs.gov/pdffiles1/Digitization/159775NCJRS.pdf

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied, 19*(4), 345-357. doi: 10.1037/a0034596

Duncan, M. J. (2006). A signal detection model of compound decision tasks (Tech. Rep. No. No. TR2006 –256). Toronto, ON: Defence Research and Development Canada.

Ebbinghaus, H. E. (1885). Memory: A contribution to experimental psychology. New York: Dover.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Eyewitness Misidentification (2019). *Innocence Project*. Retrieved from https://www.innocenceproject.org/causes/eyewitness-misidentification/

Fitzgerald, R. J., Oriet, C., & Price, H. L. (2015). Suspect filler similarity in eyewitness lineups: A literature review and a novel methodology. *Law and Human Behavior*, *39*(1), 62-107. doi: 10.1037/lhb0000095

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*(2), 151-164. doi: 10.1037/a0030618

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *15*(2), 140-152. doi:10.1037/a0015082

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *1*(4), 221-228. doi: 10.1016/j.jarmac.2012.09.003

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107-128. doi:10.3102/10769986006002107

Horry, R., & Brewer, N. (2016). How target–lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General*, *145*(12), 1615-1634. doi: 10.1037/xge0000227

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556-564. doi: 10.1037/0021-9010.70.3.556

Luus, C. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior*, *15*(1), 43-57. doi: 10.1007/BF01044829

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. New York: Cambridge University Press.

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*(4), 361-376. doi: 10.1037/a0030609

Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(2), 124-138. doi: 10.1037/0097-7403.33.2.124

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*(3), 247-255. doi: 10.1037/h0093923

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16*(4), 387-398. doi: 10.1037/a0021034

Reynolds, A., Garton, R., Kvam, P., Griffin, V. Sauer, J.D., Osth, A. and Heathcote, A. (2019) A multiple-threshold race model of deciding not to decide. *Manuscript in preparation.*

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Eyewitness identification. In N. Brewer & A. B. Douglass (Eds.), *Psychological science and the law* (pp. 208-237). New York, NY: Guildford Press.

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves the applied value of lineups without improving memory performance: Commentary on Colloff, Wade, and Strange (2016). *Psychological Science, 29*(9), 1548-1551. doi: 10.1177/09567717698528

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. (2001). Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison. *Law and Human Behavior, 25*(5), 459-473. Retrieved from http://www.jstor.org/stable /1394515

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*(1), 99-139. doi: 10.1037/a0021650

Thomaz, C. E., & Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing, 28*(6), 902-913. doi: 10.1016/j.imavis.2009.11.005

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied, 8*(3), 180-193. doi: 10.1037/1076-898X.8.3.180

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 479-496.doi: 10.1016/S0022-5371(81)90129-8

Weber, N., & Perfect, T. (2013). Why telling a witness that it's OK to say they don't know is good for justice. *Jury Expert: The Art and Science of Litigation and Advocacy*, *25*(3), 1-7. Retrieved from https://heinonline.org/

Wells, G. L. (2014). Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Direction in Psychological Science, 23*(1), 11-16. doi: 10.1177/0963721413504781

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121*(2), 262-276. doi: 10.1037/a0035940

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences*, *113*(2), 304-309. doi: 10.1073/pnas.1516814112

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology*, *105*, 81-114. doi: 10.1016/j.cogpsych.2018.06.001

Zarkadi, T., Wade, K. A., & Stewart, N. (2009). Creating fair lineups for suspects with distinctive features. *Psychological Science*, *20*(12), 1448-1453. doi: 10.1111/j.1467- 9280.2009.02463.x

Appendix A

Table 2. Discrimination Results in Easy and Hard Trials

	Easy			Hard		
Condition	*M*	*SD*	95% CIs	*M*	*SD*	95% CIs
Y/N	.80	.50	[.57, .104]	.70	.47	[.49, .92]
2AFC	1.04	.49	[.81, 1.27]	.58	.34	[.42, .74]
2AOC	1.29	.48	[1.06, 1.51]	.86	.29	[.72, .99]

Table 3.

Overall SDT-CD Results for 2AOC

Participant			Observed			Model				
	d'	*c*	CID	FID	FA	CID	FID	FA	G_{total}	*p*
1	1.54	-.06	.56	.10	.23	.57	.09	.24	.22	.974
2	.68	-.31	.44	.30	.41	.51	.23	.49	8.23	.041
3	.92	-.34	.52	.20	.44	.53	.18	.45	.36	.949
4	1.05	-.78	.65	.19	.60	.65	.19	.60	.01	.999
5	1.73	-.59	.74	.09	.39	.74	.09	.39	.002	.999
6	1.66	-.01	.56	.08	.20	.57	.08	.21	.12	.990
7	1.64	-.47	.69	.12	.34	.71	.10	.36	1.31	.726
8	.96	-.55	.58	.21	.51	.59	.20	.53	.39	.943
9	.755	-.21	.46	.17	.46	.44	.19	.44	.85	.837
10	.77	-.40	.51	.20	.52	.50	.21	.51	.17	.982
11	.95	-.60	.58	.23	.52	.61	.20	.55	1.62	.654
12	.97	-.17	.49	.18	.36	.50	.16	.38	.46	.928
13	1.14	-.19	.52	.17	.32	.55	.15	.35	2.07	.558
14	.94	.02	.42	.20	.27	.46	.16	.31	4.33	.228
15	1.80	-.50	.60	.06	.20	.59	.07	.20	.05	.997
16	.41	-.02	.33	.22	.42	.33	.21	.43	.08	.995
17	.92	-.52	.56	.23	.50	.59	.20	.52	1.89	.596
18	1.48	-.34	.63	.13	.33	.65	.11	.35	.56	.905
19	1.17	-.27	.55	.14	.38	.55	.14	.38	.01	.999
20	.49	-.40	.43	.29	.53	.46	.26	.56	1.60	.659

Notes. CID = correct identifications from target-present lineups; FID = filler identifications from target-present lineups; FA = filler identifications from target-absent lineups.

Table 4. SDT-CD Results for 2AOC, Easy Trials

Participant			Observed			Model				
	d'	*c*	CID	FID	FA	CID	FID	FA	G_{total}	*p*
1	1.70	-.004	.57	.10	.18	.59	.080	.20	.89	.827
2	.80	-.28	.46	.30	.34	.54	.22	.45	8.08	.044
3	1.07	-.36	.56	.17	.43	.57	.16	.43	.07	.995
4	1.42	-.81	.73	.10	.58	.70	.13	.54	1.73	.629
5	1.98	-.68	.79	.05	.40	.78	.07	.38	.72	.870
6	2.34	-.23	.72	.02	.20	.70	.04	.17	1.71	.634
7	1.97	-.37	.72	.09	.25	.74	.07	.27	.76	.858
8	1.21	-.55	.63	.17	.46	.65	.16	.48	.28	.965
9	.78	-.25	.48	.14	.49	.44	.18	.45	1.92	.589
10	.91	-.30	.51	.18	.43	.51	.18	.44	.02	.999
11	1.22	-.57	.64	.14	.49	.63	.15	.48	.011	.991
12	1.14	-.26	.54	.16	.37	.55	.15	.38	.15	.986
13	1.49	-.11	.57	.11	.25	.57	.10	.26	.10	.992
14	1.22	.07	.46	.09	.27	.44	.11	.25	.60	.896
15	2.46	-.28	.75	.03	.18	.75	.03	.17	.07	.995
16	.34	-.02	.31	.24	.42	.33	.23	.44	.41	.939
17	1.13	-.43	.59	.19	.43	.61	.16	.45	.54	.909
18	1.88	-.52	.74	.10	.30	.77	.08	.34	1.40	.706
19	1.45	-.39	.64	.10	.38	.63	.11	.37	.12	.989
20	.68	-.36	.47	.23	.49	.48	.22	.51	.17	.982

Table 5. SDT-CD Results for 2AOC, Hard Trials

Participant			Observed			Model				
	d'	*c*	CID	FID	FA	CID	FID	FA	G_{total}	*p*
1	1.37	-.12	.55	.10	.30	.54	.11	.29	.08	.994
2	.56	-.35	.43	.30	.48	.47	.25	.53	1.63	.653
3	.78	-.31	.48	.22	.45	.50	.20	.47	.318	.957
4	.69	-.72	.54	.30	.62	.58	.26	.65	1.32	.724
5	1.47	-.49	.67	.14	.38	.69	.12	.40	.65	.885
6	1.15	.15	.42	.15	.21	.45	.12	.24	1.16	.761
7	1.39	-.56	.67	.15	.43	.69	.13	.45	.52	.916
8	.70	-.53	.52	.25	.56	.53	.24	.57	.18	.980
9	.74	-.17	.44	.19	.42	.44	.19	.42	.01	.999
10	.65	-.51	.51	.21	.60	.49	.23	.57	.53	.913
11	.72	-.64	.53	.31	.55	.58	.26	.61	3.62	.306
12	.81	-.12	.43	.19	.36	.45	.18	.38	.36	.949
13	.85	-.25	.47	.25	.38	.52	.20	.42	2.98	.395
14	.71	-.04	.37	.30	.28	.46	.21	.38	10.32	.016
15	1.21	.15	.44	.10	.23	.43	.11	.23	.02	.999
16	.48	-.01	.35	.19	.42	.34	.20	.41	.08	.994
17	.73	-.62	.53	.27	.57	.56	.25	.60	.98	.807
18	1.05	-.16	.50	.15	.36	.50	.15	.56	.01	.999
19	.92	-.17	.47	.12	.38	.48	.17	.38	.04	.998
20	.32	-.43	.39	.35	.56	.44	.30	.61	2.23	.525

Appendix B

UNIVERSITY of TASMANIA

14 June 2019

Dr Jim Sauer
C/- University of Tasmania

Sent via email

Dear Dr Sauer

REF NO: H0018178
TITLE: Similarity Effects on Recognition: Testing the Diagnostic Feature-Detection Hypothesis

We are pleased to advise that acting on a mandate from the Tasmania Social Sciences HREC, the Chair of the committee considered and approved the above project on 12 June 2019.

Please ensure that all investigators involved with this project have cited the approved versions of the documents listed within this letter and use only these versions in conducting this research project.

This approval constitutes ethical clearance by the Tasmania Social Sciences HREC. The decision and authority to commence the associated research may be dependent on factors beyond the remit of the ethics review process. For example, your research may need ethics clearance from other organisations or review by your research governance coordinator or Head of Department. It is your responsibility to find out if the approvals of other bodies or authorities are required. It is recommended that the proposed research should not commence until you have satisfied these requirements.

In accordance with the National Statement on Ethical Conduct in Human Research, it is the responsibility of institutions and researchers to be aware of both general and specific legal requirements, wherever relevant. If researchers are uncertain they should seek legal advice to confirm that their proposed research is in compliant with the relevant laws. University of Tasmania researchers may seek legal advice from Legal Services at the University.

All committees operating under the Human Research Ethics Committee (Tasmania) Network are registered and required to comply with the *National Statement on the Ethical Conduct in Human Research* (NHMRC 2007 updated 2018).

Therefore, the Chief Investigator's responsibility is to ensure that:

(1) All investigators are aware of the terms of approval, and that the research is conducted in compliance with the HREC approved protocol or project description.

(2) Modifications to the protocol do not proceed until **approval** is obtained in writing from the HREC. This includes, but is not limited to, amendments that:

Human Research Ethics Committee (Tasmania) Network
Research Ethics and Integrity Unit
Office of Research Services
Private Bag 1
Hobart Tasmania
7001
Australia
T +61 3 6226 6254
E ss.ethics@utas.edu.au
ABN 30 764 374 782 /CRICOS 00586B
utas.edu.au

UNIVERSITY of TASMANIA

(i) are proposed or undertaken in order to eliminate immediate risks to participants;
(ii) may increase the risks to participants;
(iii) significantly affect the conduct of the research; or
(iv) involve changes to investigator involvement with the project.

Please note that all requests for changes to approved documents must include a version number and date when submitted for review by the HREC.

(3) Reports are provided to the HREC on the progress of the research and any safety reports or monitoring requirements as indicated in NHMRC guidance. Researchers should notify the HREC immediately of any serious or unexpected adverse effects on participants.

(4) The HREC is informed as soon as possible of any new safety information, from other published or unpublished research, that may have an impact on the continued ethical acceptability of the research or that may indicate the need for modification of the project.

(5) All research participants must be provided with the current Participant Information Sheet and Consent Form, unless otherwise approved by the Committee.

(6) This study has approval for four years contingent upon annual review. A *Progress Report* is to be provided on the anniversary date of your approval. Your first report is due 12 June 2020, and you will be sent a courtesy reminder closer to this due date. Ethical approval for this project will lapse if a Progress Report is not submitted in the time frame provided

(7) A *Final Report* and a copy of the published material, either in full or abstract, must be provided at the end of the project.

(8) The HREC is advised of any complaints received or ethical issues that arise during the course of the project.

(9) The HREC is advised promptly of the emergence of circumstances where a court, law enforcement agency or regulator seeks to compel the release of findings or results. Researchers must develop a strategy for addressing this and seek advice from the HREC.

Should you have any queries please do not hesitate to contact me on (03) 6226 2975 or via email ss.ethics@utas.edu.au.

Yours sincerely

Jude Vienna-Hallam
Executive Officer I Social Sciences

Human Research Ethics Committee (Tasmania) Network
Research Ethics and Integrity Unit
Office of Research Services

Private Bag 1
Hobart Tasmania
7001
Australia

T +61 3 6226 6254
E ss.ethics@utas.edu.au
ABN 30 764 374 782 /CRICOS 00586B
utas.edu.au

Appendix C

UNIVERSITY of TASMANIA

FACULTY OF HEALTH

School of Medicine

Facial Recognition Study
PARTICIPANT INFORMATION SHEET

Research team Jim Sauer, Division of Psychology, UTAS
Andrew Heathcote, Division of Psychology, UTAS
Matt Palmer, Division of Psychology, UTAS
Ellie Newton, Honours Student, Psychology, UTAS

1. Invitation

You are invited to participate in a psychology research study examining memory. This study is being conducted in partial fulfilment of an Honours degree for Miss Newton under the supervision of Dr Sauer.

2. What is the purpose of this study?

This research is being conducted as a research project within the University of Tasmania Health Sciences (Psychology) as an honours project. The aim is to investigate factors that may influence the accuracy of recognition memory and how this can be enhanced.

3. How is the study being funded?

This research is being funded by the Honours research fund.

4. Why have I been invited to participate?

You may have been invited for a couple of reasons. First, you may have signed up via Psychology's First Year participation pool. Second, you may have contacted us after seeing advertisements around campus. Finally, we may have contacted you because you have previously added your name to our list of people interested in paid research opportunities.

You are eligible to take part in this study because you are over 18 years and have normal or corrected-to-normal eyesight (you wear glasses/contact lenses). Your participation is voluntary, and there are no consequences should you decide not to participate (e.g. it will not affect your marks).

5. What will I be asked to do?

Participating in this study will take approximately 60 minutes. In this study, you will complete a number of blocks of trials. Each block will contain a study phase in which you will view a series of faces; and a test phase in which your memory will be tested. At the end of each block you can have a break.

6. Are there any possible benefits from participation in this study?

We do not expect there to be any direct benefits for participants in this study, other than a warm fuzzy feeling because you did your bit for science. However, the information gained from this study will advance our knowledge of factors that affect recognition memory and will be applicable to important real-world scenarios like

eyewitness lineups. Participating in this study is a valuable way to contribute to this important area of research.

You will be reimbursed for your time with either one hour of research credit or a $30 gift voucher.

7. Are there any possible risks from participation in this study?
There are no foreseeable risks related to participating in this study.

8. What if I change my mind during or after the study?
That is fine – you are free to withdraw at any time during the study. However, as data are stored anonymously, once you have completed the study, we will be unable to withdraw your data.

9. What will happen to the data when this study is over?
All data will be securely stored on a University of Tasmania server for five years from the date of thesis completion. This server is password-protected and only the researchers will have access. After five years the data will be archived. This means that de-identified data may be archived in an online data repository and could be used again in the future. This is in line with best practice open science recommendations.

The data will be anonymous and non-identifiable. We will not collect any identifying information from you.

10. How will the results of the study be published?
Results of the study will be published in a bound thesis held within the Division of Psychology and as a poster presentation. The results may also be published in an academic journal or as the subject of posters or presentations at conferences.

You will not be identifiable in any publication of results.

11. What if I have questions about this study?
If you have any queries please feel free to contact the research by email: newtonek@utas.edu.au or the Research Supervisor: jim.sauer@utas.edu.au

This study has been approved by the Tasmania Health and Medical/Social Sciences Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study, you can contact the Executive Officer of the HREC (Tasmania) Network on (03) 6226 6254 or email human.ethics@utas.edu.au ss.ethics@utas.edu.au. The Executive Officer is the person nominate to receive complaints from research participants. You will need to quote H0018178.

12. How can I agree to be involved?

If you would like to take part in this research, please sign the consent form.

Thank you for your time!

UNIVERSITY of TASMANIA

FACULTY OF HEALTH

School of Medicine

Facial Recognition Study
PARTICIPANT CONSENT FORM

Research team Jim Sauer, Division of Psychology, UTAS
Andrew Heathcote, Division of Psychology, UTAS
Matt Palmer, Division of Psychology, UTAS
Ellie Newton, Honours Student, Psychology, UTAS

By signing below, I confirm that I have read and understood the information sheet and in particular:

- I understand that my involvement in this research will involve approximately one hour and involve looking at images and having my memory for those images tested.
- I have been informed of and understand the purposes of this study.
- Any questions that I have asked have been answered to my satisfaction.
- I understand that my participation in this research is voluntary
- I understand that I am free to withdraw at any time, without explanation or penalty
- I understand that I will not be able to withdraw my data after completing the study as my data will be stored anonymously.
- I understand that all study data will be securely stored on the University of Tasmania's password protected server for five year. I understand that after this time the data (including my de-identified data) will be archived online and may be re-used in future research projects.
- I understand that the results of the study will be published so that I cannot be identified as a participant
- I agree to participate in the study

Please complete the form overleaf to provide your consent to participate in the research.

UNIVERSITY of TASMANIA

FACULTY OF HEALTH

School of Medicine

Participant details

Name	
Signature	
Date	

Statement by Researcher

☐ I have explained the project and the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation.

If the researcher has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

☐ The participant has received the Information Sheet where my details have been provided so participants have had the opportunity to contact me prior to consenting to participate in this project.

Name	
Signature	
Date	